How To and FAQs

Submitting Batch Jobs

Click here to see how to run **Discover** jobs.
Click here to see how to run **Gaussian** jobs.

In case of difficulty, contact the Scientific Application Support.

How do I change my default shell from C shell to Korn shell on the Alpha?

To select the Korn shell on the Alpha nciaxp:

Characters you type are **RED**.

You must type these characters exactly as shown.

1. log in to nciaxp Alpha

2. % cd

3. % chsh

Old shell: /bin/csh
New shell: /bin/ksh

4. % cp /seq/app/documents/.profile .

5. % cp /seq/app/documents/.kshrc .

6. log out of nciaxp Alpha

The next time you log in to nciaxp Alpha you will be using the Korn shell, with up-arrow command recall.

How do I interpret DSSP structural assignments?

H - 4-helix (alpha-helix)

B - residue in isolated beta-bridge

E - extended strand, participates in beta-ladder

G - 3-helix (3-10 helix)

I - 5-helix (pi helix)

T - H-bonded turn

S - bend

(In case of structural overlaps, priority is given to structure first in this list)

For details see: TABLEII in : Kabsch and Sander, Biopolymers, Vol. 22, pp. 2577-2637 (1983)

How do I use fasta to search my own database?

Example of using fasta with your own nucleotide database:

If you have a data file :

/users/group/username/nuc.seq

Create a file like: /users/group/username/data.list
with a line like:

NUC data$1A/users/group/username/data.list 1

The "1" at the end of the above line assumes your database is in Genbank format. If not then substitue the proper number from the list:

0 Pearson/FASTA (>SEQID - comment/sequence)
1 Uncompressed Genbank (LOCUS/DEFINITION/ORIGIN)
2 NBRF CODATA (ENTRY/SEQUENCE)
3 EMBL/SWISS-PROT (ID/DE/SQ)
4 Intelligenetics (;comment/SEQID/sequence)
5 NBRF/PIR VMS (>P1;SEQID/comment/sequence)
6 GCG (version 8.0) Unix Protein and DNA (compressed)
11 NCBI Blast1.3.2 format (unix only)

(If you have a peptide database then substitute a "$0" for the "$1" above.)

Run fasta with a command like (asking for the top 20 hits and alignments):

% /seq/app/fasta/fasta -Q -b 20 -d 20 -l /users/group/username/nuc.seq \ query.seq A > /tmp/username/query.fasta

This will search with the sequence in query.seq and put the output in
/tmp/username/query.fasta

How do I run Fastlink?

To run Fastlink Programs : ilink lodscore linkmap mlink

You need a pedigree ".pre" file and a parameter file.

type:

$ analysis
$ linkage
$ makeped

Makeped will convert your pedigree ".pre" file to a ".dat" file

type:

$ lcp

Answering appropiately, selecting the program you want to run.
When finished type:

$ qpedin

This will submit an nqs batch job selecting the FASTLINK version of:

ilink lodscore linkmap mlink

How can I run multiple Blast searches?

Doing multiple BLAST searches on our Alpha:

This method builds a script that is submitted to an nqs queue. It runs GCG's NetBlast repeatedly. It was created for screening of many sequences against EST and doesn't have a polished user interface yet.

Works best if you have all your query sequences (in GCG format) in a directory with names that have a .seq extension.

TYPE :

% analysis
% gcg
% ncbi
% mblast

This makes a script that you then submit to an NQS queue with:

% qsub multiblast

This default mode will search with each *.seq file in your current directory vs. EST, reporting 15 hits with alignments for each query. Each query must not contain more than 25% N and/or X, using blastn.

You can modify the request with 7 command line parameters:

mblast [1] [2] [3] [4] [5] [6] [7]
Where:
1 - directory path to query sequences
2 - extension of query files [ie. seq]
3 - database to search
4 - number of hits to report and align for each query
5 - percent (in decimal) max allowed N and/or X bases
6 - minimum length of query
7 - blast program to use

For example if I'm in my tmp directory (/tmp/gws) and have 3 queries (one.seq, two.seq, three.seq). Typing mblast is equivalent to typing:

mblast /tmp/gws seq est 15 0.25 0 blastn

When it is finsihed I will get these files in /tmp/gws:

one.n_est
two.n_est
three.n_est

Each being an individual blast search result file. If I want to see 40 hits I would type:
% mblast . . . 40

NOTE the 3 periods as place holders for the 1st 3 default parameters.

If my query files had extension .nuc (one.nuc, two.nuc, three.nuc) and I wanted to see 50 hits vs EST's I would type:
% mblast . nuc . 50

If my query files had extension .nuc (one.nuc, two.nuc, three.nuc) and I wanted to see 50 hits vs NR I would type:
% mblast . nuc nr 50

How do I run the latest version of MFold?

For those interested in RNA folding:

MFOLD 3.0 now available on nciaxp:

Startup for mfold (3000 base limit):

TYPE: analysis
TYPE: mfoldinit

TYPE: mfold { to get command line options }

Startup for mfold (10000 base limit):

TYPE: analysis
TYPE: mfoldinit10k

TYPE: mfold { to get command line options }

DETAILS:

For ksh users:
mfoldinit is an alias for '. /seq/app/mfold-3.1/exportit'
mfoldinit10k is an alias for '. /seq/app/mfold-3.1/exportit10k'

For csh users:
mfoldinit is a alias for 'source /seq/app/mfold-3.1/setenvit'
mfoldinit10k is a alias for 'source /seq/app/mfold-3.1/setenvit10k'

Source code : /seq/app/mfold-3.1/src

Documentation : /seq/app/mfold-3.1/doc
(Postscript /seq/app/mfold-3.1/doc/manual.ps)
(HTML /seq/app/mfold-3.1/doc/manual-html/index.html)
(PDF /seq/app/mfold-3.1/doc/manual.pdf)

Data file: /seq/app/mfold-3.1/dat

How do I run PAML?

PAML (Phylogenetic Analysis by Maximum Likelihood) Version 3.0a. Programs for model fitting and phylogenetic tree reconstruction using nucleotide or amino-acid sequence data.

Available Programs:

baseml: ML analysis of nucleotide sequences.

basemlg: ML analysis of nucleotide sequences (gamma rates).

codeml: ML analysis using codon subsitution, 1=coding DNA 2=protein

pamp: Parsimony-based analyses for a given tree topology

mcmctree: Bayesian estimation of phylogenies using DNA sequence data

evolver: Used to be named listtree and does miscellaneous things.

yn00: Estimates synonymous and nonsynonymous substitution rates.

Platform: Alpha

Location: /seq/app/linkage/paml

Documentation:

/seq/app/linkage/paml/doc/paml.html
/seq/app/linkage/paml/doc/paml.readme
/seq/app/linkage/paml/doc/paml.txt
/seq/app/linkage/paml/doc/pamlDOC.pdf
/seq/app/linkage/paml/doc/pamlDOC.ps

Author: Ziheng Yang
http://abacus.gene.ucl.ac.uk/software/paml.html

Command line Usage:

% analysis
% paml

How do I run prettyplot?

We have been successful in porting PrettyPlot to the Alpha (nciaxp). PrettyPlot makes graphic alignments from pileup msf files with "boxed" matching residues or bases.

HOW to USE:

1) Make a copy of your pileup msf file in the "old GCG" format with:

% oldmsf new.msf old.msf

2) Run prettyplot with:

% prettyplot -figure old.msf{*}

3) Select your type of graphics output with setplot

4) Create that type of graphics output with:

% figure prettyplot.figure

NOTE: To see all the command line options for prettyplot use:

% prettyplot –check

How do I run SeqWeb programs?

The WEB interface to the GCG programs is now available! From your browser (Netscape or Explorer) go to:

http://seqweb.ncifcrf.gov:737/

You must login in with a username and password. Your SeqWeb username is the same as your Alpha (nciaxp) username. ----> Your seqweb password is NEW (*****NOT your Alpha password). We prefer that you CALL for your SeqWeb password:

Call for your SeqWeb password, or if you have questions, contact:

Gary W. Smythers 301-846-5778 or **Bob Stephens** 301-846-5787 or The **Help Desk** 301-846-5555

Call to get your password, go to the SeqWeb page, login, and select the bottom line "Preferences" to select a new password. Complete help is available by selecting the bottom line "Help". Call 301-846-5777 for a copy of the SeqWeb Guide.

How do I automatically initialize the GCG programs?

To have gcg initialized when you login:

If you use the c-shell add the following 2 lines to your .cshrc file:

```
source /seq/app/analysis_csh
gcg
```

If you use the k-shell add the following 2 lines to your .kshrc file:

```
. /seq/app/analysis_ksh
gcg
```

Where is your local copy of PDB (Protein Data Bank)?

The following directories are nfs mounted to numerous platforms at the ABCC:

PDB Protein Data Bank

All data files are in -----> /usr/local/databases/pdb
Note & Index files are in -> /usr/local/databases/pdb/notes
Latest Newsletter in ------> /usr/local/databases/pdb/newsletter

Mirror of "divided" PDB file organization in ---> /usr/local/databases/bnl

---

I'm having problems with my keyboard backspace - delete key.

To possibly correct the backspace key problem:

If you use the c-shell add the following line to your .cshrc file:

```
setenv TERM vt340
```

If you use the k-shell add the following line to your .kshrc file:

```
export TERM=vt340
```

When are the machines available?

The computers are available 24 hours a week, 7 days a week, with the exception of scheduled downtime, system crashes and critical maintenance.

Where is documentation for various programs on the Alpha?

CLUSTALW /seq/app/clustalw/doc
CLUSTALX /seq/app/clustalxx
DOTTER /seq/app/dotter
DSSP /seq/app/structure/dssp
EPCR /seq/app/epcr
FASTA /seq/app/fasta/doc
FASTLINK /seq/app/linkage/fastlink/doc

GENEHUNTER /seq/app/gh/gh
/seq/app/gh/gh+
/seq/app/gh/gh1.3

GENSCAN /seq/app/genscan
GPFI /seq/app/GPFI
LINTRE /seq/app/linkage/lintre
MFOLD /seq/app/mfold-3.1/doc
MISMATCH /seq/app/mismatch
PAML /seq/app/linkage/paml/doc
PHYLIP /seq/app/phylip/doc

Example NQS scripts : /seq/app/phylip/scripts

PRIMER /seq/app/primer0.5
/seq/app/primer_3_6

PROFIT /seq/app/structure/ProFitV1.8/doc
PUZZLE /seq/app/PUZZLE_40/MANUAL
RHMAP /seq/app/linkage/rhmap/rhmap30/
SAGE /seq/app/linkage/sage/v31/doc
SCAN_FOR_MATCHES /seq/app/patscan
SIMLINK /seq/app/linkage/simlink
SLINK /seq/app/linkage/slink/src
SPERM /seq/app/linkage/sperm/doc
VITESSE /seq/app/linkage/vitesse
XTLSSTR /seq/app/structure/xtlsstr

What is tfsites, and how do I get more information?

We obtain the tfsites.dat file for use with findpatterns from the

NCBI ftp site:

ncbi.nlm.nih.gov in directory /repository/TFD/datasets

Or from the Institute for Transcriptional Informatics (IFTI) web site:

http://www.ifti.org

There doesn't seem to be a regular update schedule, or a release number as in the past. For more information check out the web site **http://www.ifti.org**

To read the tfsites.dat file with literature references use:

$ typedata tfsites.dat | more

If you want to check the literature reference of a particular hit, for example "E-box_CS" use:

$ typedata tfsites.dat | grep "E-box_CS"

Or:

$ grep "E-box_CS" /seq/app/gcg/gcgcore/data/moredata/tfsites.dat

---

I'm running out of diskquota and need to store some of my files on the archive. Can you give me some pointers on how to do this?

The UniTree file system provides you with a long term storage solution for your files. Each user is granted access to this facility from all platforms at the ABCC.

Access to UniTree is granted to each user in the
/archive/group/username directory.

You may change the current directory to the ARCH area by using the set default command:

% cd /archive/group/username

To return to your home directory:

% cd

Some examples are probably the best explanation:

COMMAND ACTION

% ls /archive/group/username
displays the contents of your archive area

% cat /archive/group/username/file
types out the specified file located in the archive directory

% cp file /archive/group/username
copies a file in the current directory to the archive directory

% cp /archive/group/username/file .
copies a file from the archive directory to the current directory

Unfortunately, to create a subdirectory in your ARCH area or to copy to a subdirectory of your ARCH area, you must use the FULL pathname to the directory, for instance:

% mkdir /archive/group/username/sub_dir

% cp file /archive/group/username/sub_dir

What is your GenPept database?

For details on GenPept:

anonymous ftp to **ftp.ncifcrf.gov**
cd /pub/genpept
get announce.119
get and uncompress gprel111.txt.Z

Or
take your browser to **ftp.ncifcrf.gov**
click on pub/
click on genpept/
click on announce.119
click on gprel119.txt.Z

But basically, we build GenPept from GenBank entries with FEATURES indicating a coding sequence, and extract that coding sequence to create a GenPept entry.

GenPept does not have any cross-reference data, other than back to Genbank (if Genbank entry with Locus ABCDEF has a FEATURES section indicating 3 exons, or coding regions, then there will be 3 GenPept entries ABCDEF_1, ABCDEF_2, and ABCDEF_3, one for each exon.)

GenPept is updated daily, by combining a full release and a daily file. The full release of GenPept is created when a full GenBank is released. Then we create a cumulative daily update from the GenBank cumulative daily update in the updates directory as : gpseq_updates.dat.Z

See the release text file, gprel119.txt, section 4.2 for citation info:

If you have used GenPept in your research, please include a reference to the database in all publications related to that research. For instance:

1. GenPept (GenBank Gene Products) Database. Distributed on the Internet via anonymous FTP from ftp.ncifcrf.gov, under the auspices of the National Cancer Institute's Advanced Biomedical Computing Center.